# The Role of the Statistician in Oncology Research Teams

Elizabeth Garrett-Mayer, Ph.D.
Associate Professor
Hollings Cancer Center
Dept. of Biostatistics, Bioinformatics, and Epidemiology

# Statistics

- Statistics is the art/science of **summarizing data** and **quantifying evidence**
- Better yet…summarizing data so that non-statisticians can understand it
- Scientific investigations usually involve collecting a lot of data.
- But, at the end of your study, what you really want is a "punch-line:"
  - Did the new treatment work?
  - Are the two groups being compared the same or different?
  - Is the new method more precise than the old method?
- Statistical inference is the answer!

# Do you need a statistician as part of your research team?

- YES!
- Simplest reasons:  s/he will help to optimize
  - Design
  - Analysis
  - Interpretation of results
  - Conclusions

# What if I already know how to calculate sample size and perform a t-test?

- Statisticians might know a better approach
- Trained more formally in design options
- More “bang for your buck”
- Tend to be less biased
- Adds credibility to your grant application
- Use resources that are available to you

# Different Roles

- Very collaborative
  - Active co-investigator
  - Helps develop aims and design
  - Brought in early in planning
  - Continues to input throughout trial planning and while study continues
- Consultants
  - Inactive co-investigator
  - Often not brought in until:
    - You need a sample size calculation several days before submission
    - Study/paper has been criticized/rejected for lack of statistical input
    - You've collected all of the data and don't know what to do next.
  - Only involved sparsely for planning or for analysis.

# Find a statistician early

- Your study can only benefit from inclusion of a statistician
- Statisticians cannot always rescue a poorly designed study after it has begun.
- "Statistical adjustment" in analysis does not always work.
- Ignorance is not bliss:
  - Some investigators are trained in statistics
  - But usually not all aspects!
  - Despite inclination to choose a particular design or analysis method, there might be better ways.

# Statisticians: Specific Responsibilities

- Design
  - Choose most efficient design
  - Consider all aims of the study
  - Particular designs that might be useful
    - Cross-over
    - Pre-post
    - Factorial
  - Sample size considerations
  - Interim monitoring plan

# Statisticians: Specific Responsibilities

- Assistance in endpoint selection
  - Subjective vs. objective
  - Measurement issues
    - Is there measurement error that should be considered?
    - Is the outcome actually an average of triplicate measures?
  - Multiple endpoints (e.g. tumor shrinkage AND time to death)
  - Clinical benefit versus biologic/PK endpoint
  - Continuous versus categorical outcomes

# Statisticians: Specific Responsibilities

- Analysis Plan
  - Statistical method for **EACH** aim
  - Account for type I and type II errors
  - Stratifications or adjustments are included if necessary
  - Simpler is often better
  - Loss to follow-up:  plan for missing data

## Most common problems seen in study proposals when a statistician is not involved

- Outcomes are not clearly defined
- There is not an analysis plan for secondary aims of the study
- Sample size calculation is too simplistic or absent
- Assumptions of statistical methods are not appropriate

# Four examples today

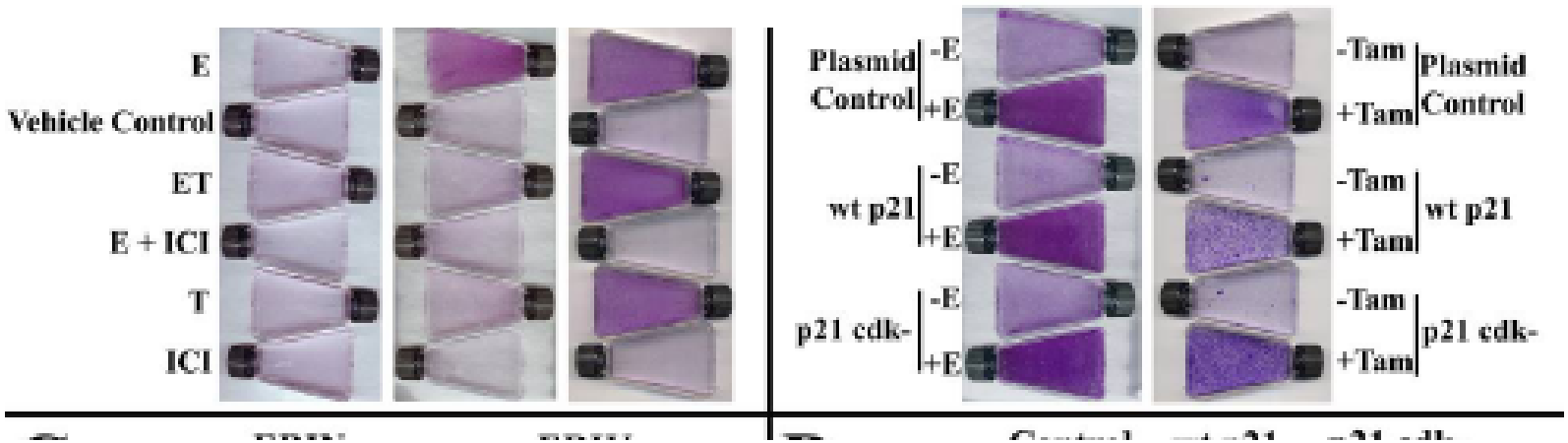
1. Statistical analysis to quantify differences
2. Statistical analysis to identify biomarkers
3. Sample size calculations to determine how many mice to study
4. Taking your agents to the clinic: phase I study design

# Four examples today

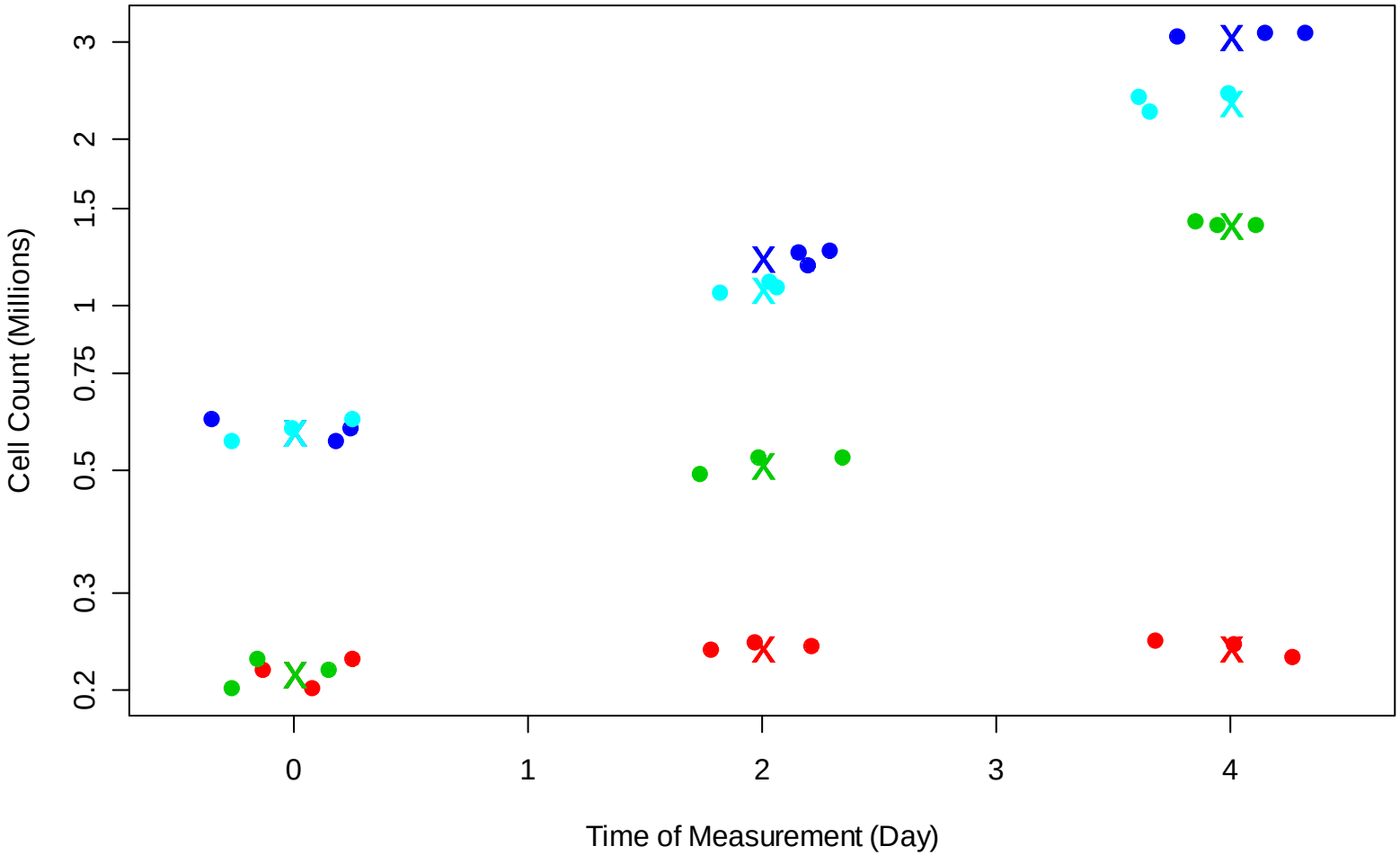
1. Statistical analysis to quantify differences
2. Statistical analysis to identify biomarkers
3. Sample size calculations to determine how many mice to study
4. Taking your agents to the clinic: phase I study design

# Ben Ho Park et al. NEJM brief report (to appear): "Tamoxifen stimulated growth of breast cancer due to loss of p21 expression"

- "Some resistant breast cancers develop a growth proliferative response to [tamoxifen], as evidenced by tumor regression upon its withdrawal. We present here a patient whose breast cancer displayed tamoxifen growth stimulation concurrent with loss of the cyclin dependent kinase inhibitor p21. Our study demonstrates that loss of p21 expression in conjunction with Tamoxifen exposure leads to aberrant [ER] alpha phosphorylation and a subsequent growth proliferative response."
- "Obvious association"
- But…..
  - What IS the level of evidence?
  - Reviewers require quantification



# Here is some of the data:



# How to quantify?

- Comparisons of interest
  - Are there changes over time
    - Ratio of day 2 to day 0 (within lines)
    - Ratio of day 4 to day 0 (within lines)
  - Differences in ratios comparing cell lines
    - Day 2 to day 0 ratio comparing lines
    - Day 4 to day 0 ratio comparing lines
- Noted complications
  - There are multiple pieces of data that go into creating each ratio
  - How do we calculate the standard errors to make these comparisons?

# Hmmmm….i don't think a t-test will do.

- Solution:  Linear regression model
- Throw all the data into one big analytic model
  - "borrows strength"
  - Improves efficiency
  - Allows us to make all comparisons
  - Use all of the data instead of comparing summaries

# Model

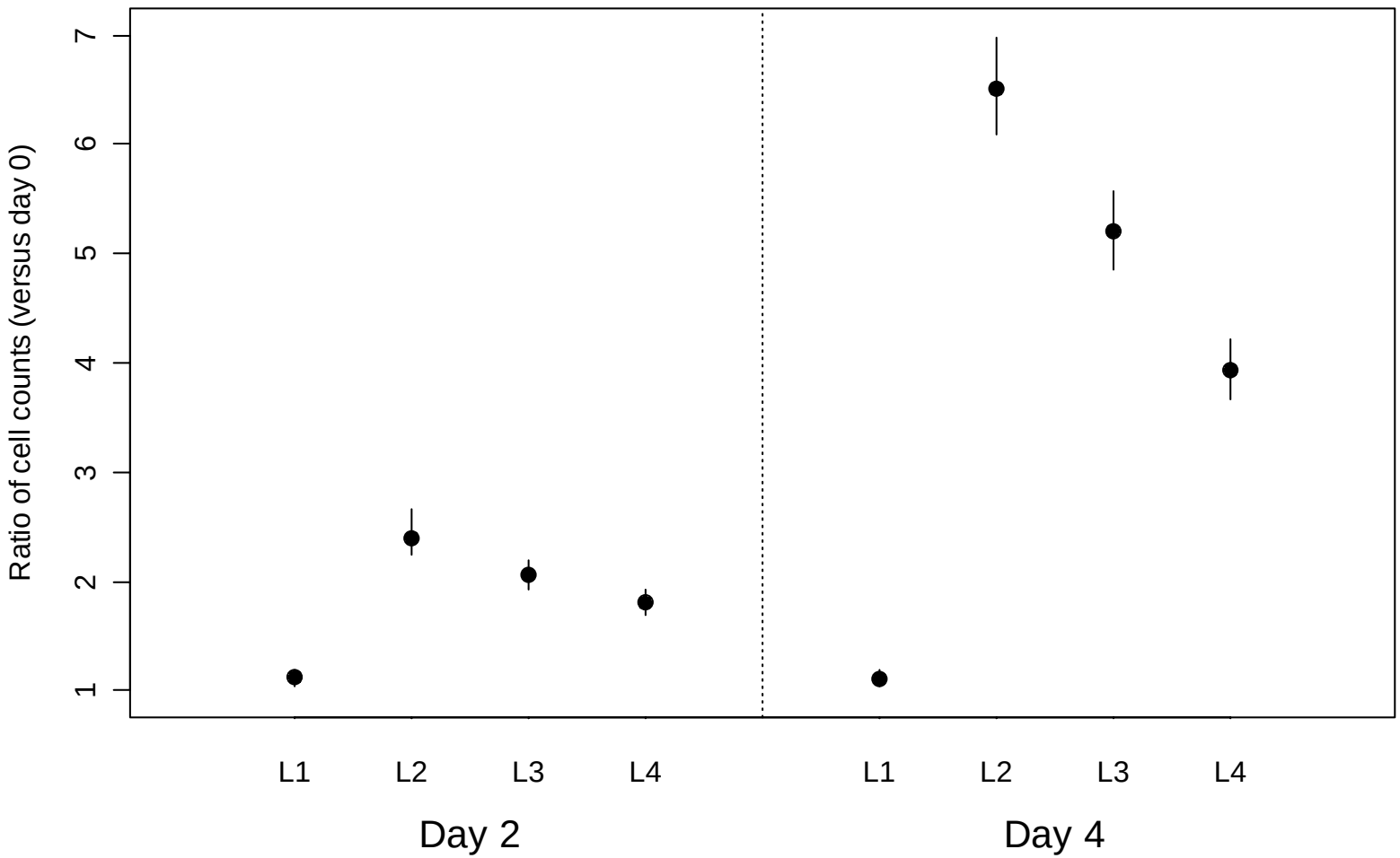
$$\log(cellcount) = \beta_0 + \beta_1 day2 + \beta_2 day4 + \beta_3 line2 + \beta_4 line3 + \beta_5 line4 + \beta_6 day2:line2 + \beta_7 day2:line3 + \beta_8 day2:line4 + \beta_9 day4:line2 + \beta_{10} day4:line3 + \beta_{11} day4:line4$$

- Of note:
  - 12 parameters (β's) to estimate
  - 4 cell lines x 3 days = 12 combinations

# How does this model help us?

- Example 1:
  - Is the day 2 to day 0 ratio for cell line 2 different than 1?
  - $\beta_1 + \beta_6 = 0$ ?
- Example 2:
  - Are the day 2 to day 0 ratios different for cell lines 2 and 3?
  - $\beta_7 - \beta_6 = 0$ ?
- Example 3:
  - Are the day 4 to day 0 ratios different than the day 2 to day 0 ratios for cell line 4?
  - $\beta_2 + \beta_{11} - \beta_1 - \beta_8 = 0$ ?

# New figure


Ratio of cell counts (versus day 0)
1
2
3
4
5
6
7
L1
L2
L3
L4
L1
L2
L3
L4
Day 2
Day 4

# New Table

| Cell line | Day | Estimated Ratio | p-value | 95% CI |
|---|---|---|---|---|
| 1 | Day 2 to day 0 | 1.12 | 0.002 | 1.04, 1.19 |
| 2 | Day 2 to day 0 | 2.40 | 0.003 | 2.24, 2.66 |
| 3 | Day 2 to day 0 | 2.06 | <0.001 | 1.92, 2.20 |
| 4 | Day 2 to day 0 | 1.81 | <0.001 | 1.69, 1.93 |
| | | | | |
| 1 | Day 4 to day 0 | 1.11 | <0.001 | 1.04, 1.19 |
| 2 | Day 4 to day 0 | 6.51 | <0.001 | 6.09, 6.97 |
| 3 | Day 4 to day 0 | 5.20 | <0.001 | 4.85, 5.58 |
| 4 | Day 4 to day 0 | 3.94 | <0.001 | 3.67, 4.22 |

# Importance of this application

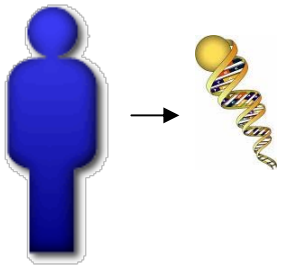
- Very common design!
- This approach can be used in many studies

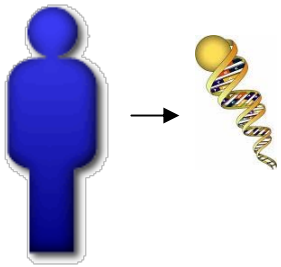
# Four examples today

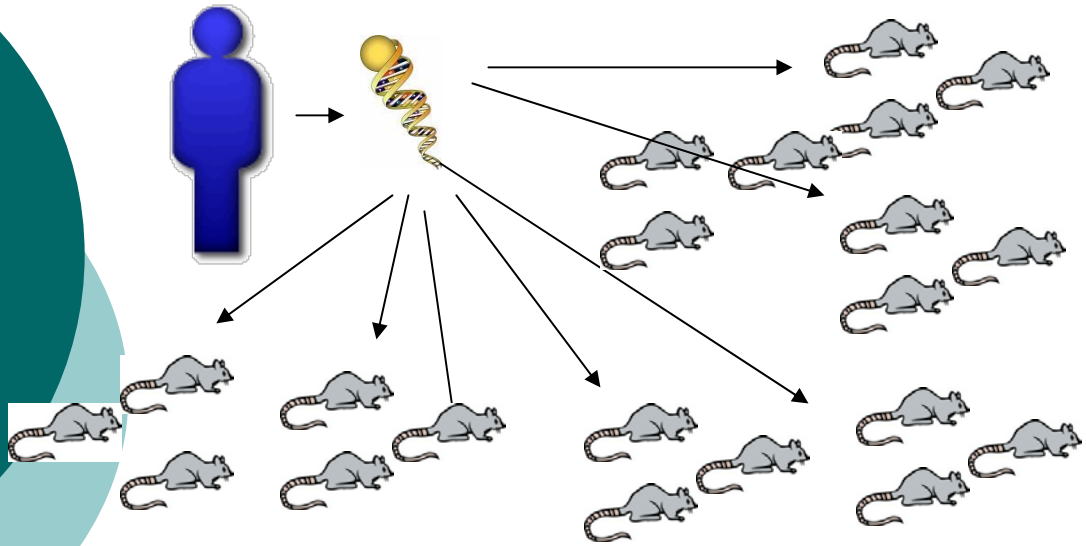
1. Statistical analysis to quantify differences
2. Statistical analysis to identify biomarkers
3. Sample size calculations to determine how many mice to study
4. Taking your agents to the clinic: phase I study design

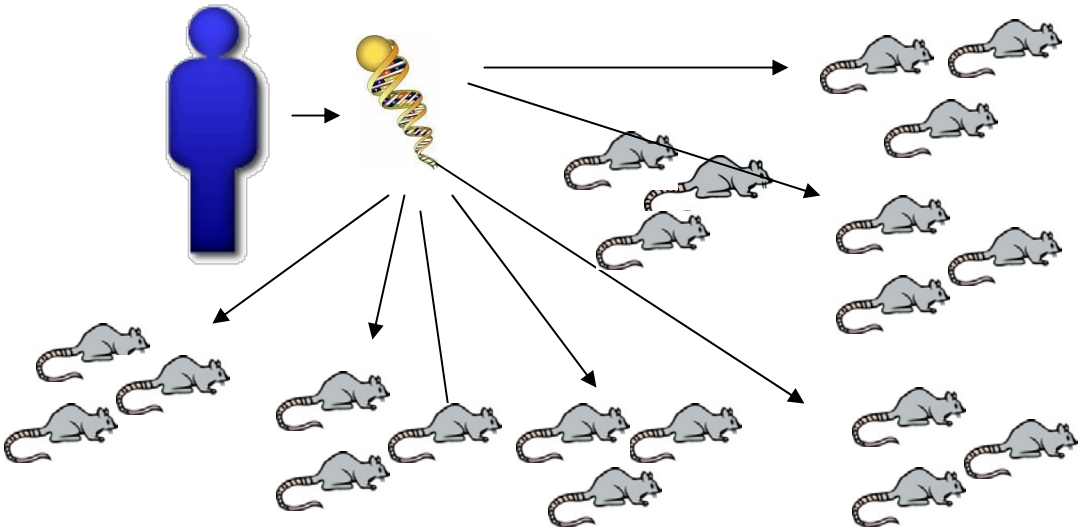
# Example 2: Identifying Biomarkers in Pancreatic Cancer Samples (Jimeno, Hidalgo, et al.)

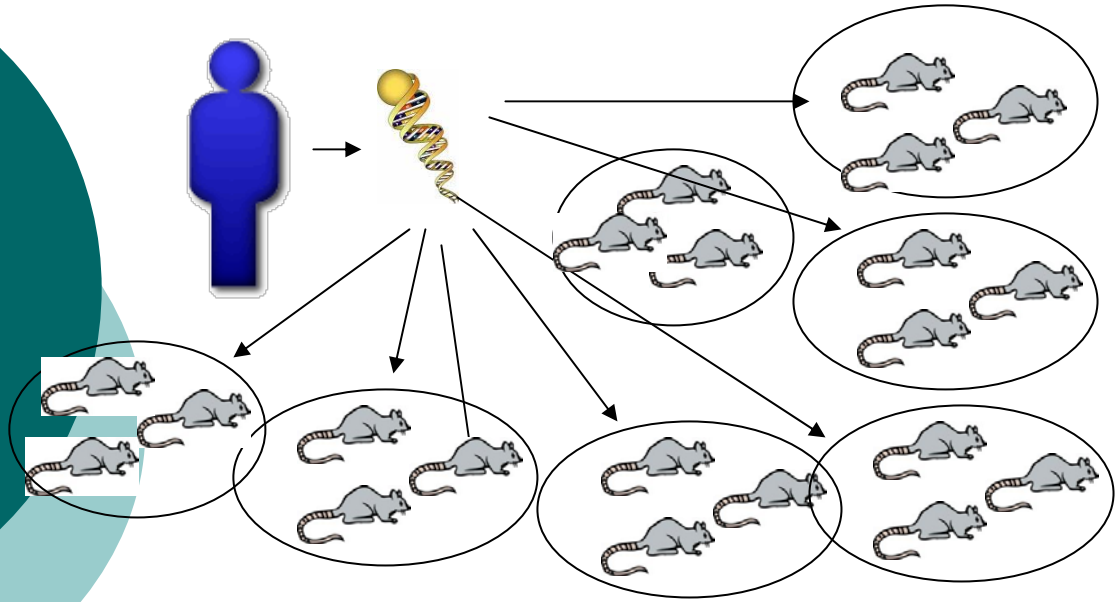
- Collaboration between Hopkins and MD Anderson quantitating proteins in pancreatic cancer tumors
- 7 cases (samples)
- 7 treatment settings
  - Control, CCI-779, OSI-774, CI-1040, and 2 way combinations.
  - CCI-779 = rapamycin (mTOR inhibitor)
  - OSI-774 = tarceva (EGFR inhibitor)
  - CI-1040 = MEK inhibitor
- Hence, 49 conditions (7 cases x 7 treatments)
- 1-3 mice (xenografts) per condition

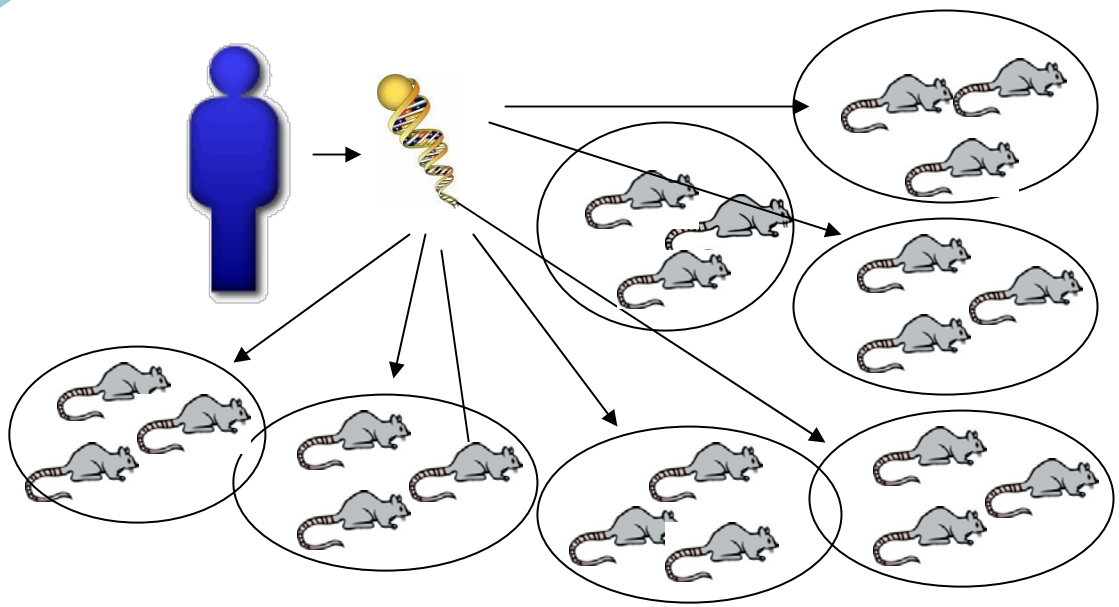


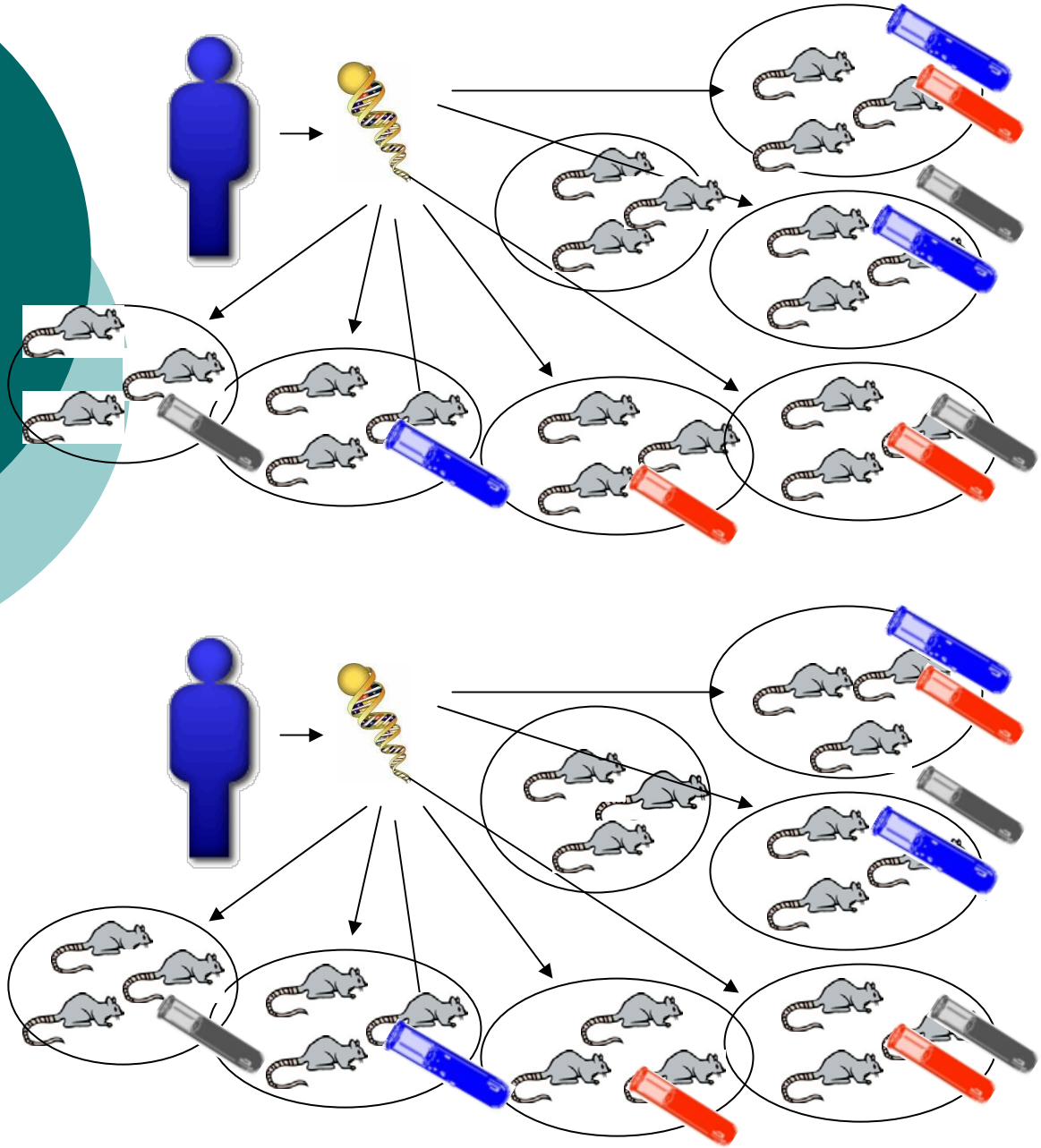












= rapamycin

= tarceva

= CI-1040

# Example 2: Identifying Biomarkers in Pancreatic Cancer Samples

- Sampled tumors from xenografts
- **Measured 49 different proteins**
- Outcomes of interest:
  - Association between control levels and efficacy of treatment
  - Association between changes in proteins (due to treatments) and efficacy
  - **Differences in protein expression under different treatment conditions**

# Statistical Challenges

- 49 markers
- 7 cases x 7 treatment conditions x 1-3 mice
- A lot of data! >7000 data points
- **How to pick out which markers are inhibited in which conditions?**
  - Individual analyses will have low power and will lead to multiple comparisons problems
  - Joint model will improve power and avoid multiple comparisons issues.
  - Recall that all of this data came from tumors grown from just SEVEN cases!

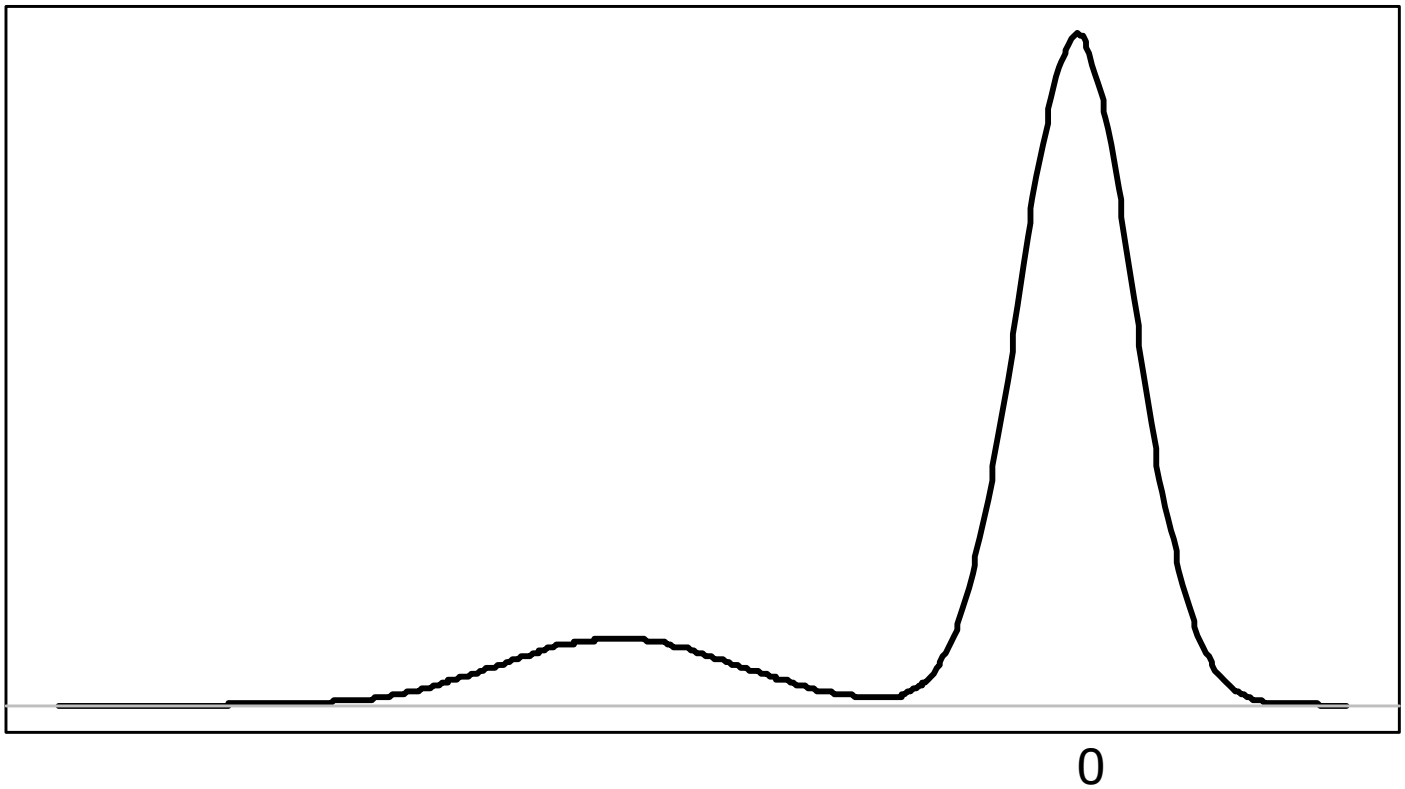
# Naïve Approach: Individual analyses

- **Standardize by the control mice**
  - divide by expression in control mice
  - take log
  - value of 0 implies no difference between treatment and control
  - **Now, only 6 treatment conditions to consider**
- **Look at each of the 49x6 = 294 treatment marker combinations**
- **Determine whether or not there is evidence that standardized mean is different than 0.**
- **Use simple t-test (294 times!)**
- **Problems: low power and multiple comparisons**

# Joint model

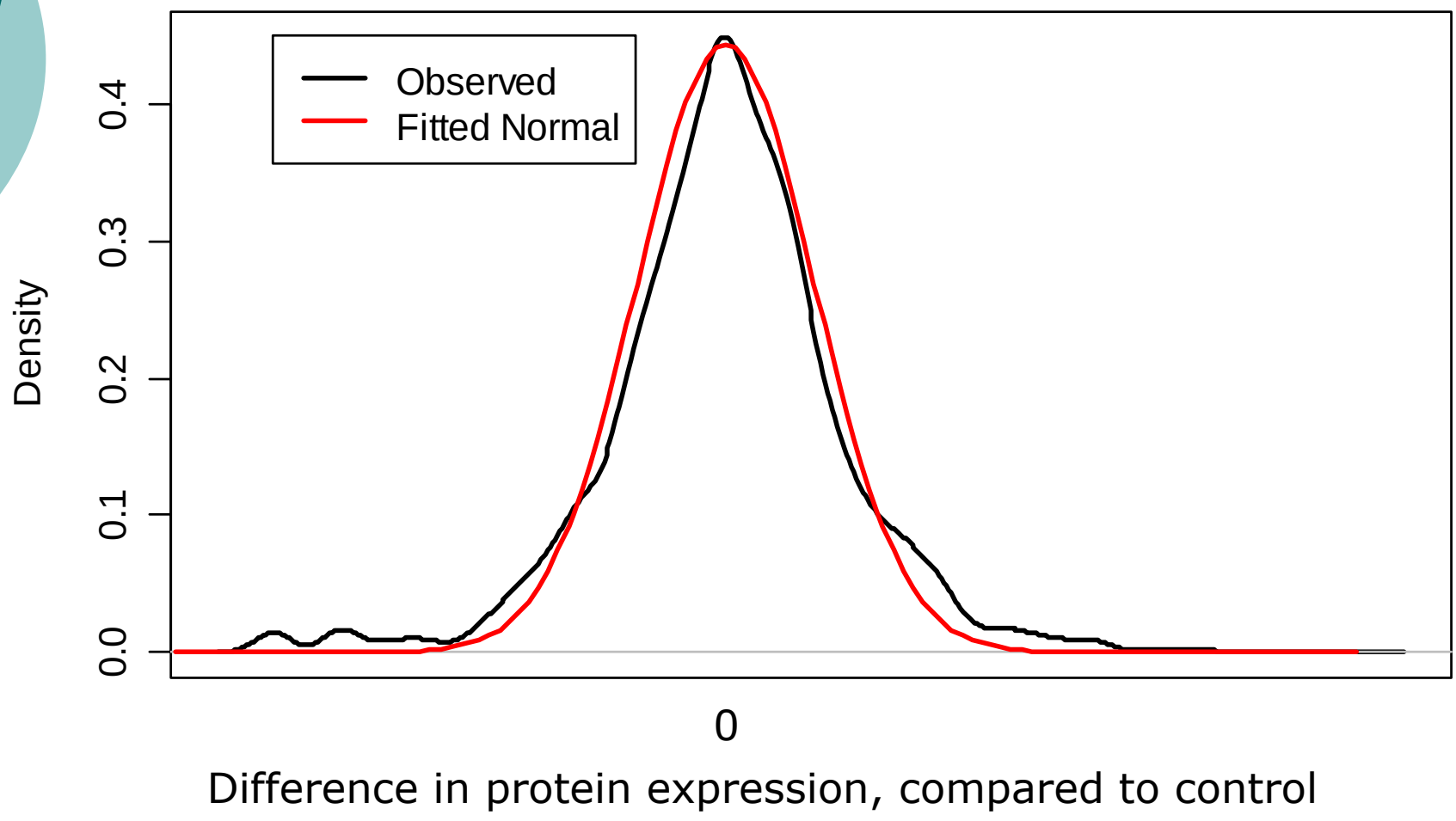
- Assumption: There are two categories for a marker to belong to for each treatment
  - Inhibited (under-expressed as compared to control)
  - Normally expressed (as compared to control)
- Normally expressed: acknowledges that there is some variation in "normally" expressed.
- Each of these categories can represented by a normal (Gaussian) distribution.
- Instead of p-values, we would like to know
  - **to which category each treatment-marker combination belongs**
  - **how strong the evidence of categorization is**

# Assumed model (extreme (ideal) case)




Difference in protein expression, compared to control

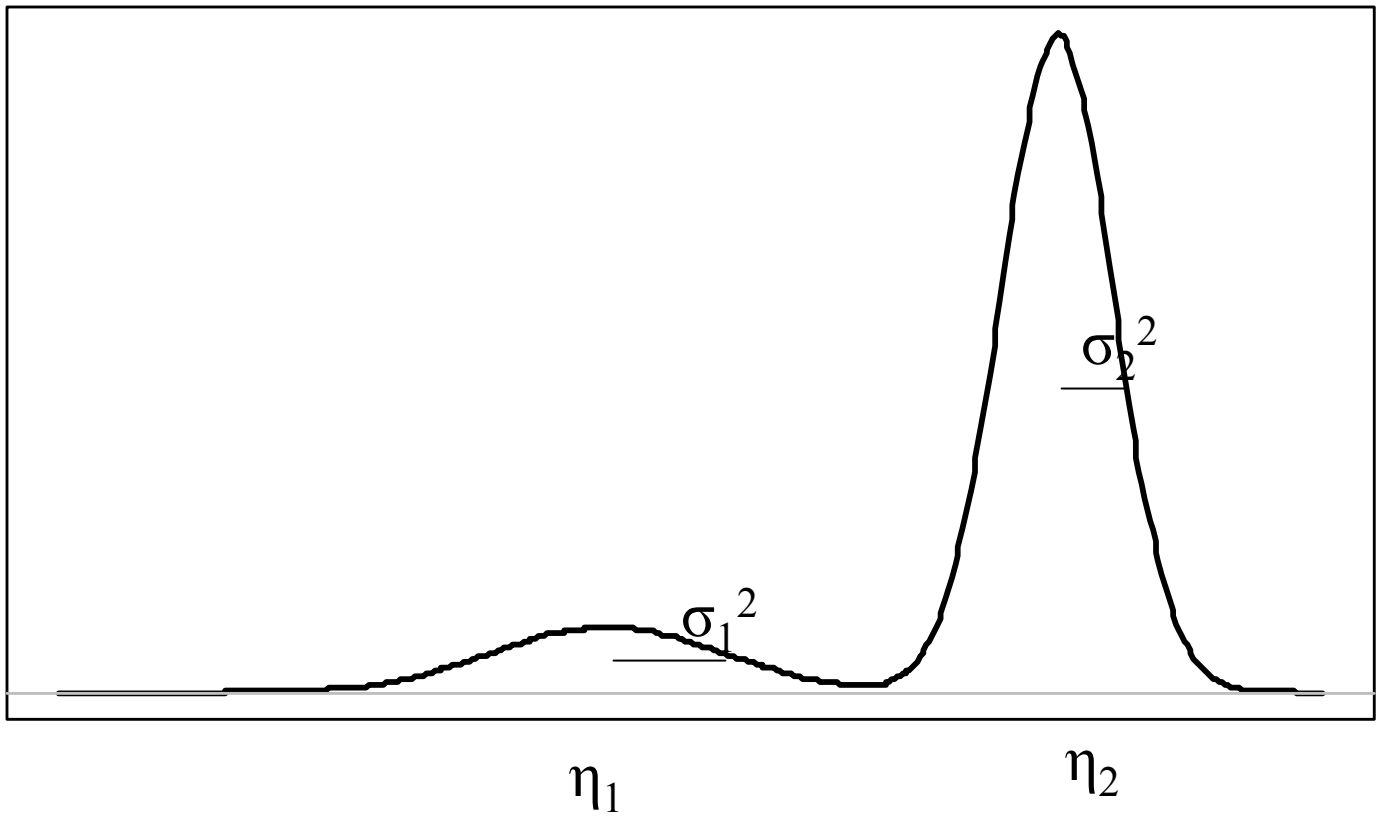
# Observed Data



# Empirical Evidence

- Seems reasonable model to fit
- Based on empirical data, relatively few in underexpressed component
- Adjusts for "case":  multiple mice have tumors from same case
- **treatment x marker effects are parameters of interest**

# Model



# Model

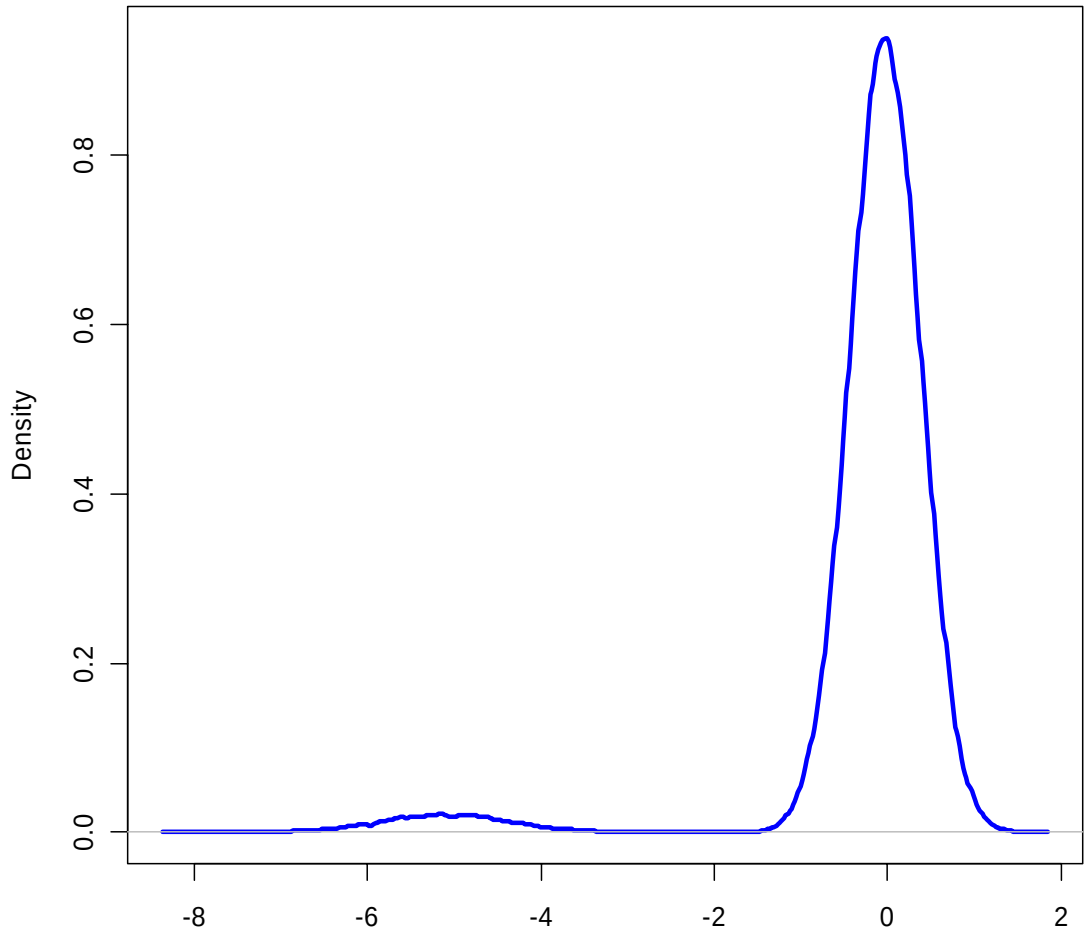
- $y_{ijkm}$=log(trt/ctrl)
- i = mouse, j = case, k = trt, m = marker
  - i = 1,…,≈150
  - j = 1,…,7
  - k = 1,…,6 (adjusted for control)
  - m = 1,…,49

$$(1)\ \log(y_{ijkm}) = \alpha_i + \beta_j + \lambda_{km}$$

$$(2)\ \lambda_{km} \sim N(\eta_1, \sigma_1^2)\ \text{w.p.}\ \pi_1$$

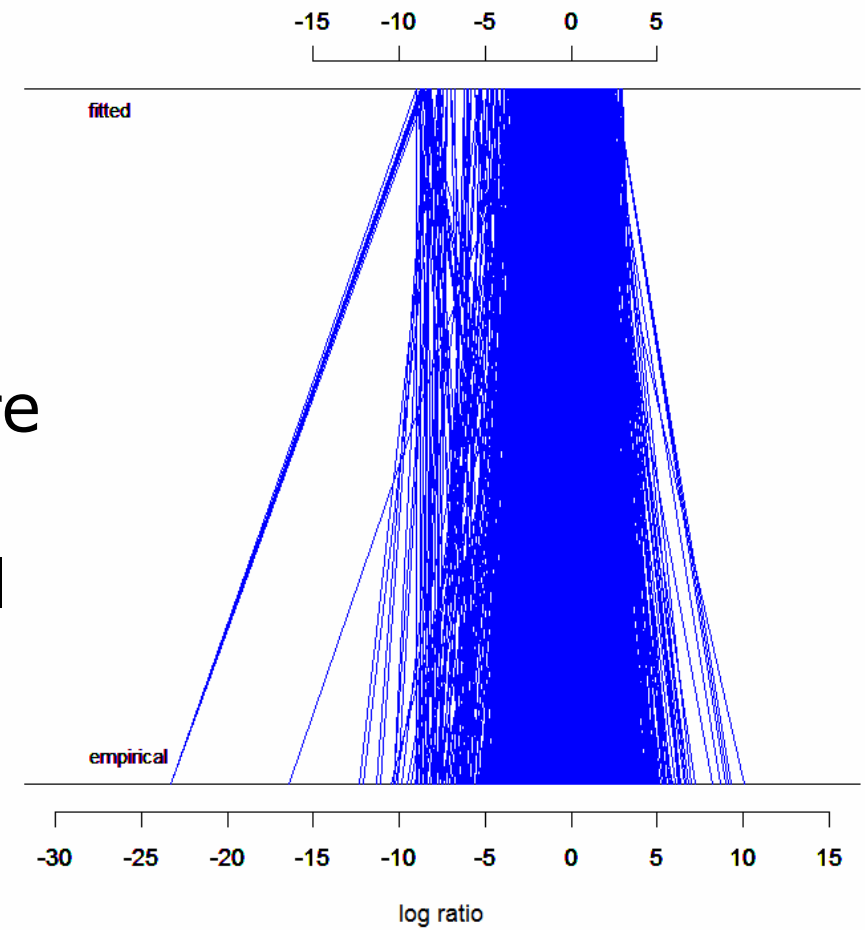
$$\sim N(\eta_2, \sigma_2^2)\ \text{w.p.}\ \pi_2$$

# Fitted Model

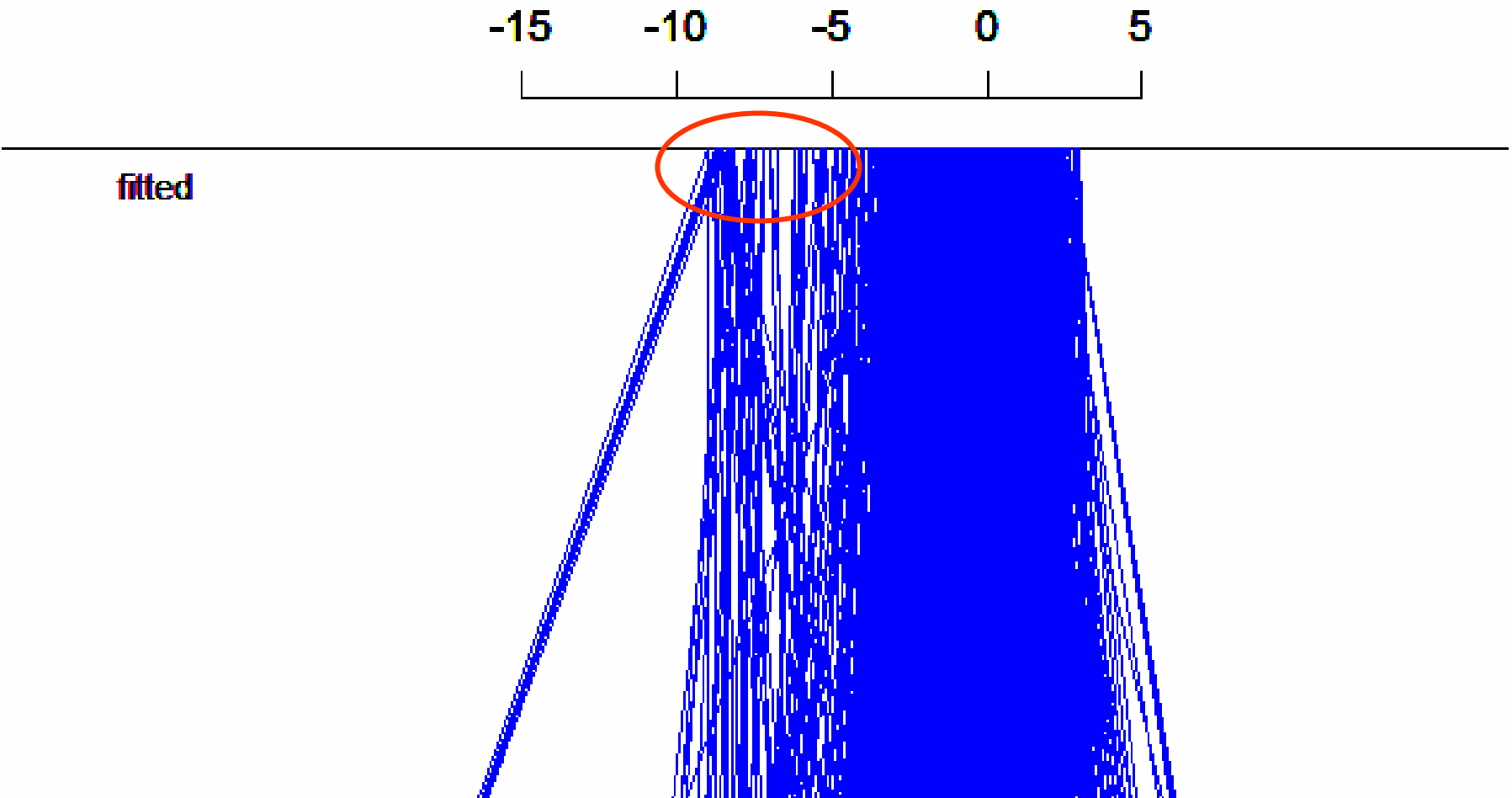


# Results

- "shrinkage" of ≈7000 values (49 markers x 150 mice = 7350)
- some of the extreme values are imprecise
- hierarchical model uses information about all genes to better understand variance



# Inhibited Marker Values



# Inferences

- Inferences are based on strength of evidence for which component a marker-treatment combination belongs to
- Very clear-cut results here: posterior probabilities were very close to 0 or 1.
- No p-values!
- No type I error!
- No multiple comparisons!

# Markers showing inhibition

| Marker/RX | OSI-774 | CCI-779 | CI-1040 | CI-1040+CCI-779 | OSI-774+CCI-779 | OSI-774+CI-1040 |
|---|---|---|---|---|---|---|
| Akt | 1 | 1 | 1 | 1 | 1 | 1 |
| b.cat | 1 | 1 | 1 | 1 | 1 | 1 |
| b3.Int | 1 | 1 | 1 | 1 | 1 | 1 |
| Bcl2 | 1 | 1 | 1 | 1 | 1 | 1 |
| c.Jun | 1 | 1 | 1 | 1 | 1 | 1 |
| c.Kit | 1 | 1 | 1 | 1 | 1 | 1 |
| CDK2 | 1 | 1 | 1 | 1 | 1 | 1 |
| CDK4 | 1 | 1 | 1 | 1 | 1 | 1 |
| coll.Au | 1 | 1 | 1 | 1 | 1 | 1 |
| COX2 | 1 | 1 | 1 | 1 | 1 | 1 |
| E.cad | 1 | 1 | 1 | 1 | 1 | 1 |
| EGFR | 1 | 1 | 1 | 1 | 1 | 1 |
| ERK2 | 1 | 1 | 1 | 1 | 1 | 1 |
| GSK3 | 1 | 1 | 1 | 1 | 1 | 1 |
| HER3 | 1 | 1 | 1 | 1 | 1 | 1 |
| IkBa | 1 | 1 | 1 | 1 | 1 | 1 |
| JNK | 1 | 1 | 1 | 1 | 1 | 1 |
| MEK1 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.Akt.308 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.Akt.473 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.BAD | 1 | 1 | 1 | 1 | 1 | 1 |
| P.c.Jun | 1 | 1 | 1 | 1 | 1 | 1 |
| P.Elk1 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.GSK3 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.HER2 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.IkBa | 1 | 1 | 1 | 1 | 1 | 1 |
| P.p38 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.p53 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.p53.1 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.p70S6K | 1 | 1 | 1 | 1 | 1 | 1 |
| P.PKCa | 1 | 1 | 1 | 1 | 1 | 1 |
| P.pRb | 1 | 1 | 1 | 1 | 1 | 1 |
| P.S6.235 | 1 | 2 | 1 | 2 | 2 | 1 |
| P.S6.240 | 1 | 2 | 1 | 2 | 2 | 1 |
| P.Src.416 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.STAT1 | 1 | 1 | 1 | 1 | 1 | 1 |
| P.STAT6 | 1 | 1 | 1 | 1 | 1 | 1 |
| p38 | 1 | 1 | 1 | 1 | 1 | 1 |
| p53 | 1 | 1 | 1 | 1 | 1 | 1 |
| PARP | 1 | 1 | 1 | 1 | 1 | 1 |
| PI3Ka | 1 | 1 | 1 | 1 | 1 | 1 |
| PI3Kd | 1 | 1 | 1 | 1 | 1 | 1 |
| PKCa | 1 | 1 | 1 | 1 | 1 | 1 |
| pRb | 1 | 1 | 1 | 1 | 1 | 1 |
| PSTAT3.705 | 1 | 1 | 1 | 1 | 1 | 1 |
| PTEN | 1 | 1 | 1 | 1 | 1 | 1 |
| S6 | 1 | 1 | 1 | 1 | 1 | 1 |
| Src | 1 | 1 | 1 | 1 | 1 | 1 |
| STAT3 | 1 | 1 | 1 | 1 | 1 | 1 |

# Results

- Two markers show inhibition
- Inhibition in these markers is observed in all treatment combinations involving rapamycin
- These markers were consistent with the preconceived hypotheses of investigators
- Statistical approach
  - Evidence based on naïve approach would have been suspect based on 294 t-tests.
  - Avoidance of p-values for inference makes conclusions simpler.

# Four examples today

1. Statistical analysis to quantify differences
2. Statistical analysis to identify biomarkers
3. Sample size calculations to determine how many mice to study
4. Taking your agents to the clinic: phase I study design

# Sample Size and Power

- The most common reason statisticians get contacted
- Sample size is contingent on design, analysis plan, and outcome
- With the wrong sample size, you will either
  - Not be able to make conclusions because the study is “underpowered”
  - Waste time and money because your study is larger than it needed to be to answer the question of interest

# Sample Size and Power

- And, with wrong sample size, you might have problems interpreting your result:
  - Did I not find a significant result because the treatment does not work, or because my sample size is too small?
  - Did the treatment REALLY work, or is the effect I saw too small to warrant further consideration of this treatment?
  - This is an issue of SCIENTIFIC versus STATISTICAL signficance

# Sample Size and Power

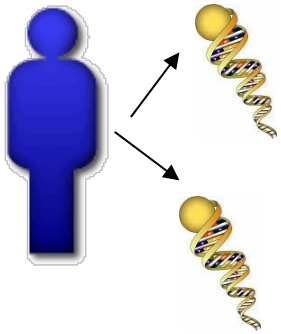
- Sample size ALWAYS requires the investigator to make some assumptions
  - How much better do you expect the experimental therapy group to perform than the standard therapy groups?
  - How much variability do we expect in measurements?
  - What would be a clinically relevant improvement?
- The statistician CANNOT tell you what these numbers should be (unless you provide data)
- It is the responsibility of the clinical/laboratory investigator to define these parameters

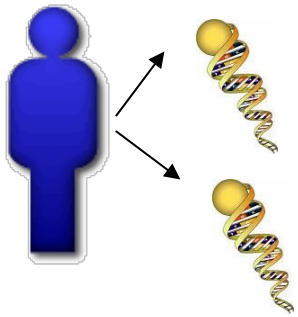
# Sample Size and Power

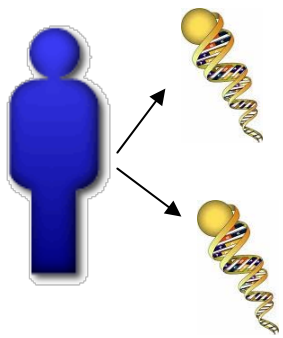
- Hypothesis testing:
  - $H_0$: new treatment does not work
  - $H_1$: new treatment works
- Review of power
  - Power = *The probability of concluding that the new treatment is effective if it truly is effective*
  - Type I error = *The probability of concluding that the new treatment is effective if it truly is NOT effective*
  - *(Type I error = alpha level of the test)*
  - *(Type II error = 1 – power)*
- **When your study is too small, it is hard to conclude that your treatment is effective**

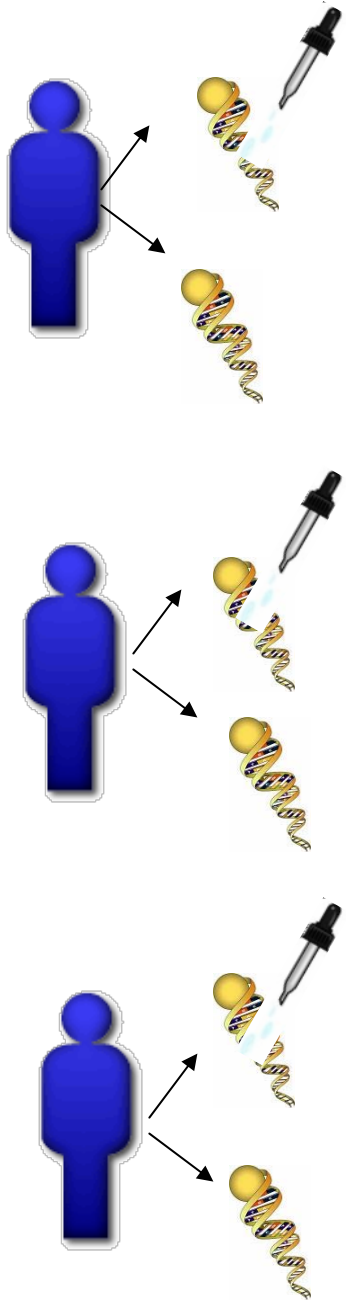
# Example 3: Cancer Stem Cells in Core Binding Factor Leukemias (Civin et al.)

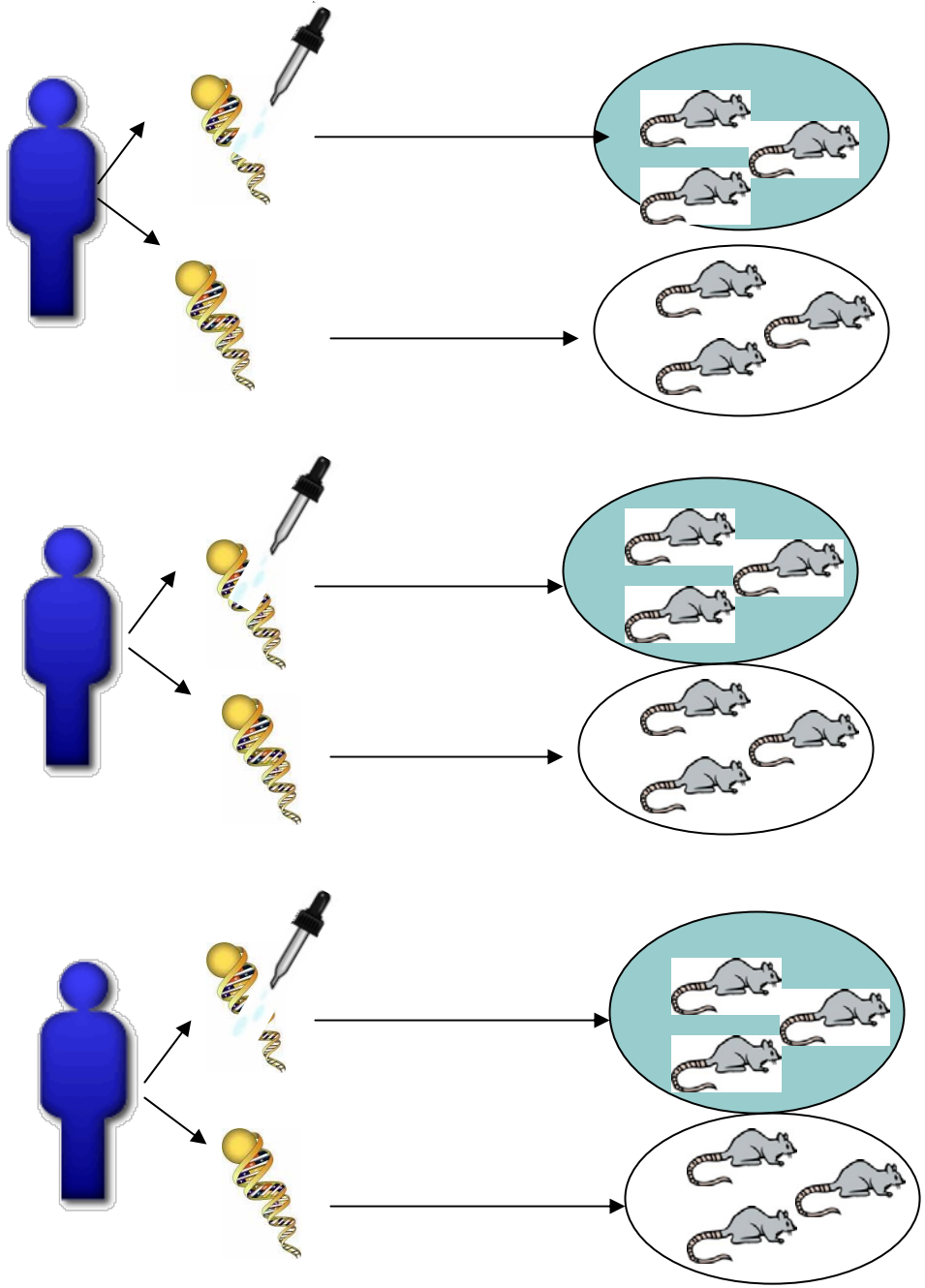
- Cancer stem cell program project grant submission
- A primary goal of project: *to determine if the engraftment frequency is different with modified AML cells compared to control AML cells*
- Study design:
  - Acquire samples from 5-20 samples (banked)
  - Use mouse xenograft approach to produce tumors
  - Transplanted cells were either modified or not (controls).
  - Expectation is that modified cells should engraft less frequently.
  - Primary modifications involved: FLT3 and KIT

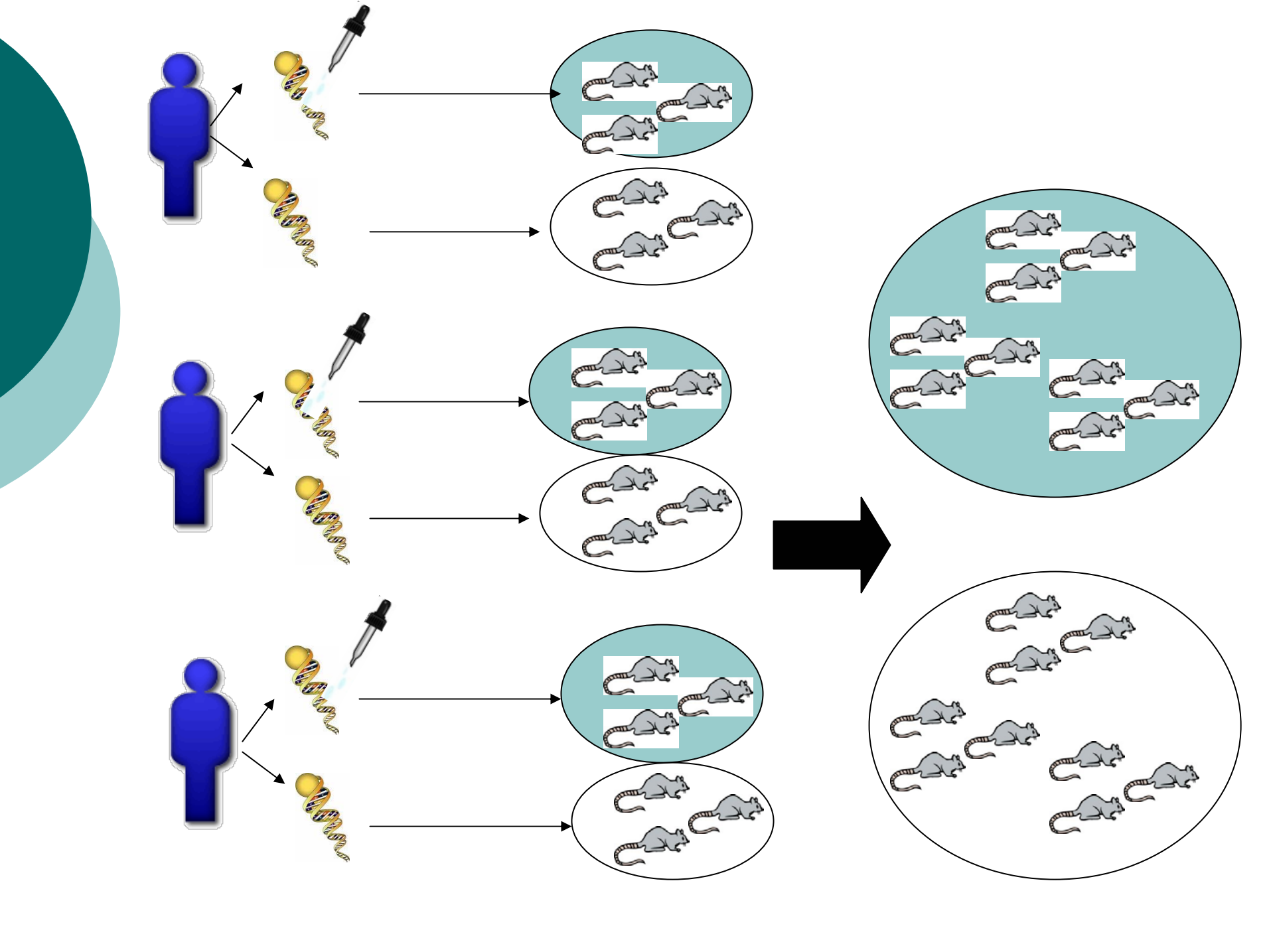












# Example 3: Cancer stem cells in Core Binding Factor leukemias

- How many mice and samples do we need?
  - We considered 3, 4, or 5 mice per condition/case
  - Up to 20 samples available
- Using samples from the same case in many mice: lose independence assumption.
- Standard issues to consider for power calculations:
  - What is expected engraftment rate in each condition?
  - What is the expected difference in engraftment rates?
  - What are the desired type I and type II error rates?
- **Novel things to consider:**
  - **What is the variability in engraftment across mice transplanted with cells from different cases?**
  - **Does variability of engraftment differ by condition?**

# Example 3: Cancer stem cells in Core Binding Factor leukemias

- Power calculations
  - Allow for variability in engraftment rates
  - Low, moderate and high degree of variability
- Perform "simulations"
  - Situation is too complex to use predetermined sample size calculations
  - Simulate data according to varying assumed data conditions
  - Determine how large N has to be to see "significant differences"
  - Usually, large number of datasets simulated (e.g., 10000)
  - Number of simulations depends primarily on time until grant submission (true!)

# Example 3: Cancer stem cells in Core Binding Factor leukemias

Table 1: Description of correlation structure of simulated datasets

| Overall expected rate of engraftment | Range* of probabilities of engraftment under LOW variability | Range* of probabilities of engraftment under MODERATE variability | Range* of probabilities of engraftment under HIGH variability |
|---|---|---|---|
| 5% | 3%, 8% | 2%, 13% | 1%, 28% |
| 10% | 6%, 15% | 4%, 23% | 1%, 45% |
| 20% | 13%, 29% | 8%, 40% | 3%, 65% |
| 30% | 21%, 41% | 14%, 54% | 5%, 76% |
| 40% | 29%, 52% | 20%, 65% | 8%, 83% |
| 50% | 38%,62% | 27%, 73% | 12%, 88% |
| 60% | 48%, 71% | 35%, 80% | 17%, 92% |
| 80% | 71%, 87% | 60%, 92% | 55%, 97% |

* range is actually the range for which 95% of the engraftment probabilities are expected to fall.

# Example 3: Cancer stem cells in Core Binding Factor leukemias

Table 2A: Minimum power to detect differences in engraftment rates between groups of mice based on experiments with N samples. Low variability under control conditions, moderate variability under modified conditions. With three mice per group and two conditions (modified and control), the number of mice is six times the number of samples. Assumes two-sided alpha of 0.10 and 3 mice per condition per case.

| Difference in engraftment rates | N=5 (30 mice) | N=8 (48 mice) | N=10 (60 mice) | N=15 (90 mice) | N=20 (120 mice) |
|---|---|---|---|---|---|
| **0.30** | 0.43 | 0.67 | 0.69 | 0.88 | 0.92 |
| **0.40** | 0.56 | 0.83 | 0.92 | 0.97 | 0.99 |
| **0.50** | 0.72 | 0.90 | 0.95 | 0.99 | >0.99 |
| **0.60** | 0.75 | 0.92 | 0.96 | >0.99 | >0.99 |
| **0.70** | 0.78 | 0.93 | 0.97 | >0.99 | >0.99 |

# Example 3: Cancer stem cells in Core Binding Factor leukemias

Table 2B: Minimum power to detect differences in engraftment rates between groups of mice based on experiments with N samples. Moderate variability under control conditions, high variability under modified conditions. With three mice per group and two conditions (modified and control), the number of mice is six times the number of samples. Assumes two-sided alpha of 0.10 and 3 mice per condition per case.

| Difference in engraftment rates | N=5 (30 mice) | N=8 (48 mice) | N=10 (60 mice) | N=15 (90 mice) | N=20 (120 mice) |
|---|---|---|---|---|---|
| **0.30** | 0.40 | 0.56 | 0.64 | 0.75 | 0.86 |
| **0.40** | 0.57 | 0.77 | 0.83 | 0.96 | 0.98 |
| **0.50** | 0.67 | 0.88 | 0.94 | 0.99 | >0.99 |
| **0.60** | 0.74 | 0.91 | 0.96 | >0.99 | >0.99 |
| **0.70** | 0.77 | 0.92 | 0.97 | >0.99 | >0.99 |

# Four examples today

1. Statistical analysis to quantify differences
2. Statistical analysis to identify biomarkers
3. Sample size calculations to determine how many mice to study
4. Taking your agents to the clinic: phase I study design

# Phase I study design

- "Standard" Phase I trials (in oncology) use what is often called the '3+3' design

Treat 3 patients at dose K

1. If 0 patients experience dose-limiting toxicity (DLT), escalate to dose K+1
2. If 2 or more patients experience DLT, de-escalate to level K-1
3. If 1 patient experiences DLT, treat 3 more patients at dose level K
   A. If 1 of 6 experiences DLT, escalate to dose level K+1
   B. If 2 or more of 6 experiences DLT, de-escalate to level K-1

- Maximum tolerated dose (MTD) is considered highest dose at which 1 or 0 out of six patients experiences DLT.
- Doses need to be pre-specified
- Confidence in MTD is usually poor.

# Should we use the “3+3”?

- **It is terribly imprecise and inaccurate in its estimate of the MTD**
- **Why?**
  - **MTD is not based on all of the data**
  - **Algorithm-based method**
- Likely outcomes:
  - Choose a dose that is too high
    - Find in phase II that agent is too toxic.
    - Abandon for further investigation or go back to phase I
  - Choose a dose that is too low
    - Find in phase II that agent is ineffective
    - Abandon agent
- **Phase I is the most critical phase of cancer drug development! USE A SMARTER DESIGN!**

# Continual Reassessment Method (CRM)

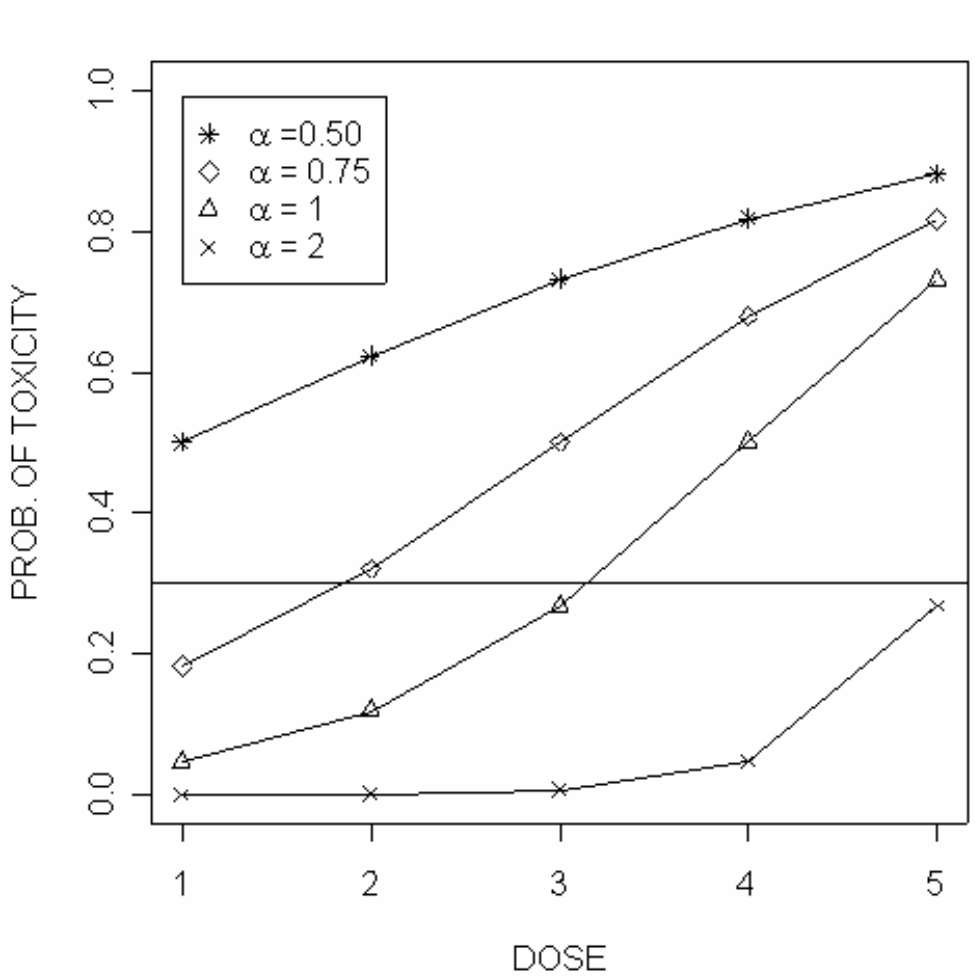
- Allows statistical modeling of optimal dose: dose-response relationship is assumed to behave in a certain way
- Can be based on "safety" or "efficacy" outcome (or both).
- Design searches for best dose given a desired toxicity or efficacy level and does so in an efficient way.
- This design REALLY requires a statistician throughout the trial.
- **ADAPTIVE**
- Example:  Phase I/II trial of Samarium 153 in High Risk Osteogenic Sarcoma (Schwartz)

# CRM history in brief

- Originally devised by O'Quigley, Pepe and Fisher (1990) **where dose for next patient was determined based on responses of patients previously treated in the trial**

- Due to safety concerns, several authors developed variants
  - Modified CRM (Goodman et al. 1995)
  - Extended CRM [2 stage] (Moller, 1995)
  - Restricted CRM (Moller, 1995)
  - and others….

# Basic Idea of CRM



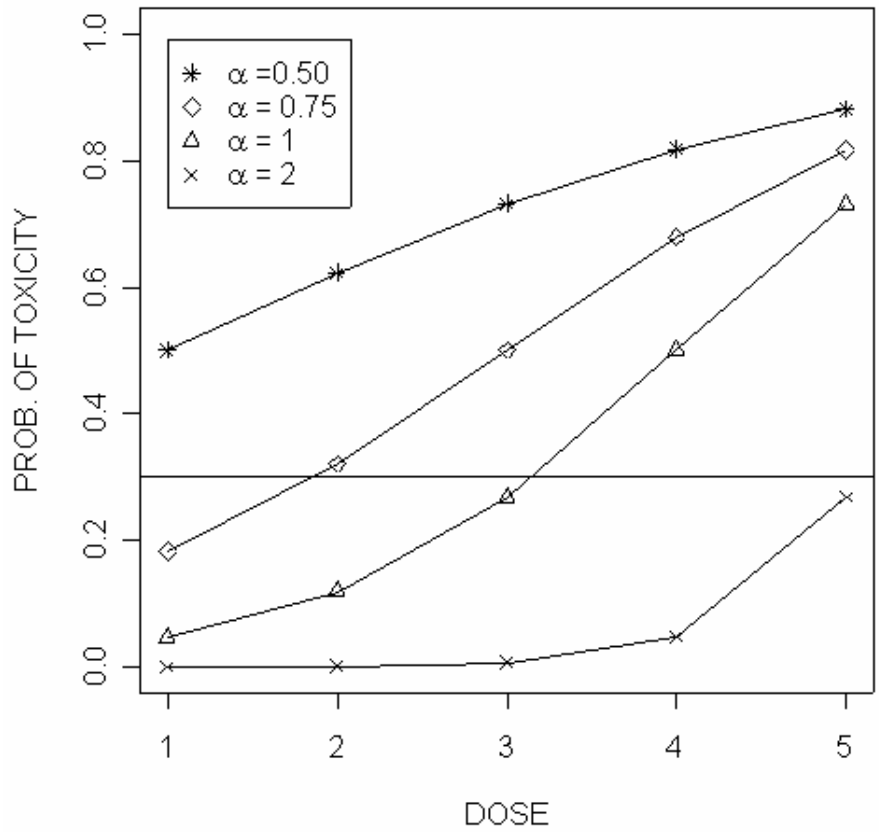

$$p(toxicity|dose = d_i) = \frac{\exp(3 + \alpha d_i)}{1 + \exp(3 + \alpha d_i)}$$

# Modified CRM

(Goodman, Zahurak, and Piantadosi, Statistics in Medicine, 1995)

Carry-overs from standard CRM

- Mathematical dose-toxicity model must be assumed
- To do this, need to think about the dose-response curve and get preliminary model.
- **We CHOOSE the level of toxicity that we desire for the MTD (e.g., p = 0.30)**
- At end of trial, we can estimate dose response curve.



## Modified CRM by Goodman, Zahurak, and Piantadosi (Statistics in Medicine, 1995)

- Modifications by Goodman et al.
  - Use 'standard' dose escalation model until first toxicity is observed:
    - Choose cohort sizes of 1, 2, or 3
    - Use standard '3+3' design (or, in this case, '2+2')
  - **Upon first toxicity, fit the dose-response model using observed data**
    - Estimate $\alpha$
    - Find dose that is closest to toxicity of 0.3.
  - Does not allow escalation to increase by more than one dose level.
  - De-escalation can occur by more than one dose level.

**Example** Samarium in pediatric osteosarcoma:

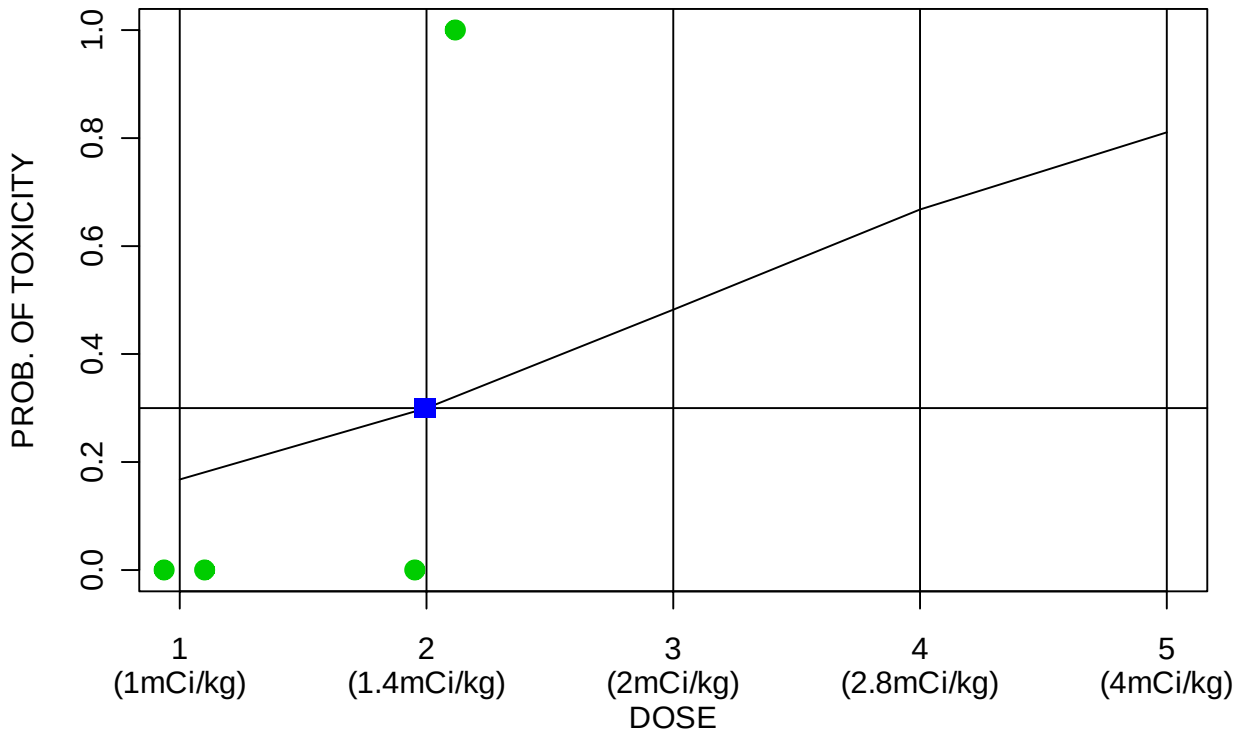
2 patients treated at dose 1 with 0 toxicities

2 patients treated at dose 2 with 1 toxicity

➔ Fit CRM using equation below

$$p(toxicity|dose = d_i) = \frac{\exp(3 + \alpha d_i)}{1 + \exp(3 + \alpha d_i)}$$

- Estimated $\alpha = 0.77$
- Estimated dose is 1.4mCi/kg for next cohort.



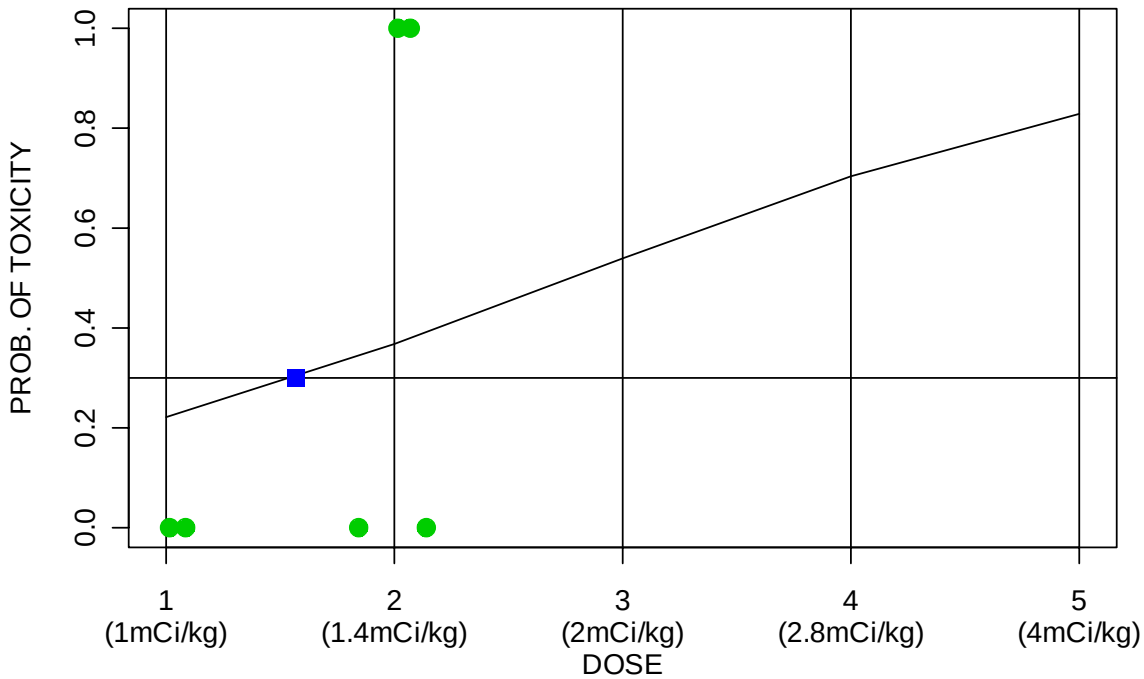
**Example** Samarium study with cohorts of size 2:

2 patients treated at 1.0 mCi/kg with no toxicities

4 patients treated at 1.4 mCi/kg with 2 toxicities

➔ Fit CRM using equation on earlier slide

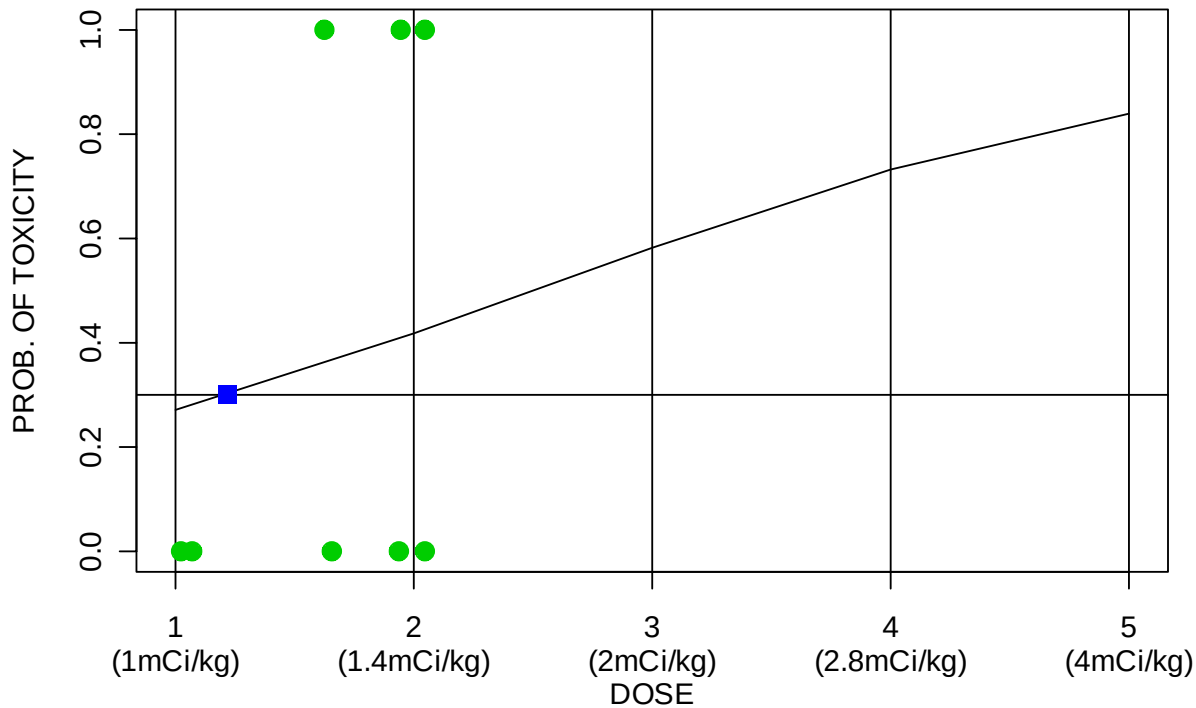
- Estimated $\alpha = 0.71$
- Estimated dose for next patient is 1.2 mCi/kg



**Example** Samarium study with cohorts of size 2:

2 patients treated at 1.0 mCi/kg with no toxicities
4 patients treated at 1.4 mCi/kg with 2 toxicities
2 patients treated at 1.2 mCi/kg with 1 toxicity
➔ Fit CRM using equation on earlier slide

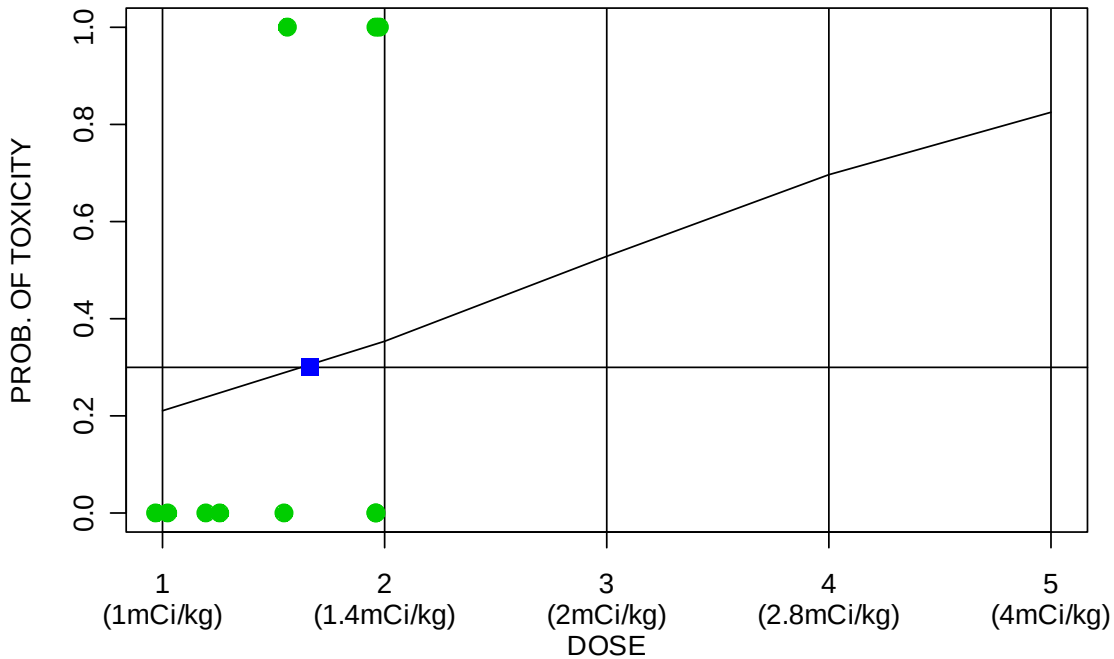
- Estimated $\alpha = 0.66$
- Estimated dose for next patient is 1.1 mCi/kg



**Example** Samarium study with cohorts of size 2:

2 patients treated at 1.0 mCi/kg with no toxicities
4 patients treated at 1.4 mCi/kg with 2 toxicities
2 patients treated at 1.2 mCi/kg with 1 toxicity
2 patients treated at 1.1 mCi/kg with no toxicities
➔ Fit CRM using equation on earlier slide

- Estimated $\alpha = 0.72$
- Estimated dose for next patient is 1.2 mCi/kg



# When does it end?

- Prespecified stopping rule
- Can be fixed sample size
- Often when a “large” number have been assigned to one dose.
- This study should enroll at least two more cohorts.

# Concluding Remarks

- **Get your statistician involved as soon as you begin to plan your study**
- Try to avoid:
  - Contacting statistician only several days before grant/protocol/proposal is due
  - Asking statisticians to rewrite rejected statistical sections
  - Asking statisticians to analyze data that have arisen from a poorly designed trial
- Statisticians have a lot to add
  - “Fresh” perspective to your study
  - Study will be more efficient!